DOCUMENT RESUME

ED 419 000 TM 028 296

AUTHOR Terwilliger, James S.
TITLE Portfolios and Classroom Assessment: Some Claims and Questions.
PUB DATE 1997-08-00
NOTE 11p.
PUB TYPE Information Analyses (070) -- Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; *Performance Based Assessment; *Portfolio Assessment; Psychometrics; Student Evaluation; Test Construction; *Test Use; Writing Evaluation
IDENTIFIERS *Authentic Assessment

ABSTRACT

This paper traces the origins of the portfolio movement, examines the treatment of portfolios in texts and other writings on classroom assessment, and poses several questions about the use of portfolios in classrooms. Interest in portfolios as assessment devices first emerged in the literature in the field of composition and writing, but interest in portfolios evolved almost concurrently in other disciplines in the language arts. The emerging literature on portfolios has had little impact upon textbooks on classroom assessment. Several claims are made by proponents of portfolio assessment, but these cannot be accepted at face value. Portfolios, claimed to be more authentic than other forms of assessment, are not necessarily so. They may not capture intended outcomes in all fields as well as they work in language arts, and they may show a bias in favor of performance over more basic outcomes such as the acquisition of knowledge. Portfolios do not offer a more psychometrically sound basis for assessment than traditional forms of assessment. They may be based on a flawed concept of growth, and they may require an inordinate investment of time and effort on the part of teachers. Although portfolios may have value as instructional aids, it is not clear that they are useful assessment devices in all subject matter fields and at all grade levels. (Contains 22 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 419 000

PORTFOLIOS AND CLASSROOM ASSESSMENT: SOME CLAIMS AND QUESTIONS

James S. Terwilliger
Department of Educational Psychology

University of Minnesota

August, 1997*

TM028276

* This paper was originally scheduled for presentation at the the 1997 AERA meetings in March. Due to a variety of family crises, completion of the paper was delayed.

BEST COPY AVAILABLE

PORTFOLIOS AND CLASSROOM ASSESSMENT: SOME CLAIMS AND QUESTIONS

INTRODUCTION

I have taught courses on classroom assessment to preservice and inservice teachers for over 30 years. For much of that period the topics in my course outlines have been quite stable and, to a large tent, based upon traditional measurement concepts and practices, e.g. The Taxonomy of Educational Objectives (Bloom, 1956), validity, reliability, various paper and pencil item formats, performance and product assessments, a brief overview of standardized tests, and the assignment of grades to students. With luck, I feel I can do justice to these topics in the 10-week term system under which I have taught.

The "school reform" movement has clearly presented a challenge to the conventional isdom under which I have operated. The new emphasis upon alternative assessment procedures has forced me to re-examine the relative emphasis I place on more traditional approaches and certain of the newer practices that stress more complex performance measures.

I have no arguement with the claim that teachers need to employ performance measures with greater frequency than has often been suggested in standard texts on classroom measurement. In fact, the authors of such texts seem to recognize this also. Newer editions tend to contain expanded treatments of performance and product evaluation procedures. The titles of the books more commonly employ the term "assessment" rather than "measurement" to reflect the broader concern for a variety of data gathering techniques.

The call for the widespread adoption of portfolios in the classroom is closely related to the emphasis upon performance measures. In fact, the use of portfolios is typically justified by the need to obtain samples of "authentic " student work. For example, a portfolio has been defined as " a purposeful collection of student work that exhibits the student's efforts, progress, and acheivement in one or more areas ." (Paulson, Paulson, and Meyer, 1991, p. 60). Valencia (1990) states that portfolios are based upon " a philosophy that honors both the process and products of learning as well as the active participation of the teacher and the students in their own evaluation and growth." (p. 340)

This paper traces the origins of the portfolio movement, examines the treatment of portfolios in texts and other writings on classroom assessment, and poses several questions about the use of portfolios in classrooms. As will be seen, I believe the case for portfolios is quite different than the case for greater use of performance measures in general. My goal is to convince the reader to give careful thought to several basic issues associated with the use of portfolios in the classroom before jumping on the portfolio bandwagon.

THE ORIGINS OF PORTFOLIOS

Interest in portfolios as assessment devices first emerged in the literature in the field of composition and writing. For example, articles by Ballard (1992), Farr (1990, 1991), Gentile (1992), Herter (1991), Howard (1990), Krest (1990), Murphy and Smith (1990), and Yancey (1992) all address the use of portfolios in the assessment of writing at the K-12 level. In contrast, discussions of issues in the use of writing portfolios at the college level are presented in articles by Black, et. al. (1994) and Hamp-Lyons and Condon (1993).

BEST COPY AVAILABLE

Interest in portfolios evolved almost concurrently in other disciplines in the language arts, especially literacy and reading. Articles by Hansen (1992), Russavage (1992), Sunstein(1992), Tierney et. al. (1991), Valencia (1990), and Valencia and Calfee (1991) all deal with the use of portfolios in these two fields. More recent articles too numerous to mention have advocated the more general use of portfolios in virtually all subject matter fields.

Typically, articles on portfolios tend to be uncritcal endorsements of portfolios which emphasize the superiority of the approach over more conventional assessment procedures, i.e. paper and pencil tests. The general literature on portfolios appears primarily in journals aimed at teachers and administrators, e.g. The Reading Teacher, Educational Leadership, and The Phi Delta Kappan. For the most part, authors of the articles are curriculum and subject matter experts who give little or no attention to technical assessment issues such as reliability and validity.

SOURCES OF INFORMATION ON PORTFOLIOS

It should come as no surprise that the emerging literature on portfolios has, to date, had little impact upon textbooks on classroom assessment. I have examined 12 "standard" texts published or revised since 1991 and found that five do not even list "portfolios" in their subject matter index. Three texts devote 1-5 pages to the topic and the remaining four devote more than six pages to the topic.

The scant attention given to portfolios by textbook authors is likely due to a combination of the natural "lag" in the incorporation of new approaches/ideas into traditional texts and a certain amount of skepticism on the part of authors who are accustomed to more traditional approaches which stress concern for psychometric perspectives. Whatever the reason, it is clear that the instructor of a course on classroom assessment must find other sources of information if the justice is to be done to the topic of portfolios.

There are, of course, books for teachers devoted exclusively to portfolios. Two books on portfolios written for teachers that I have examined tend to focus upon the advantages of portfolios over traditional assessment procedures. Little is said about the possible disadvantages. Student Portfolios (1993), an NEA publication, consists on a series of short reports written by teachers describing portfolio projects in a variety of school settings ranging from K-12 and all areas of the curriculum. Although some reports provide examples of rating forms and scoring rubrics employed in the projects, no data are presented with respect to the amount of staff time needed to implement a portfolio system, the linkage of portfolios to more traditional forms of assessment, and empirical evidence of the level of agreement among teachers in judging portfolio contents.

Portfolio and Performance Assessment (1994) by Farr and Tone is designed specifically for teachers of the language arts. The authors stress the need for "working portfolios" as opposed to "show portfolios". They link portfolios clearly to more holistic assessments consistent with " the emerging understanding of reading and writing as meaning-construction activities". The book presents a thorough arguement for the use of portfolios in the language arts curriculum and provides very detailed guidance on the development of portfolios with many examples from actual school projects. However, no data are presented supporting the relability or validity of the resulting assessments.

Probably the most balanced and comprehensive treatment of issues encountered in the design and use of portfolios is found in the article by Arter and Spandel (1992). They cover all of the major practical and technical concerns that need to be considered in undertaking portfolio assessment. They conclude with the following comments:

When using portfolios for assessment and instruction, we need to be cautious that such assessments are developed and used properly. We can be misled by work portfolios because the content looks so right. We might not notice that the material was not generated in a way to show what students can do or that is not representative of student work,etc.

This caution is as important for classroom use of portfolios as it is for portfolios in large-scale assessment. If teachers do not understand how they can be misled by poorly conceived tasks and fuzzy criteria and how extraneous performance requirements can affect student performance, then student portfolios of work will be misleading as to what students really know and can do.

Additionally, there is the danger that if we allow users to rush into use of portfolios for instruction and assessment purposes without thinking through their asessment needs, how a portfolio fits into those needs, and what potential problems they might encounter, they could very likely be confused and disappointed when the portfolio assessment does not fulfill their expectations of "fixing" all assessment problems. We want to avoid having people rush headlong into portfolio assessment and reject it later because it didn't work. Portfolios have the potential to be too useful a part of our assessment and instrucional arsenal to allow this to happen.

SOME BASIC DEFINITIONS AND PURPOSES

There are many possible variations on the basic definition of a portfolio. The one provided by Arter and Spandel (1992) is quite general and seems to describe what is typically meant when portfolios are discussed.

> Our adaptation defines a student portfolio as a purposeful collection of student work that tells a story of the student's efforts, progress, or acheivement in (a) given area(s). This collection must include student participation in the selection of portfolio content; the guidelines for selection; the criteria for judging merit; and evidence of student self-reflection.

Valencia and Calfee (1991) present a useful typology for differentiating portfolios according to the purpose they are designed to serve. The three major types they define are:

> The showcase portfolio---a collection of the student's best or favorite work. The student selects most of the entries so that the portfolio emerges as a unique portrait of the individual over time. Self-reflection, self-evaluation, and self-selection take priority over standardization.
>
> The documentation portfolio---may include observations, checklists, anecdotal records, interviews,and classroom tests, as well as performance-based assessments. Some entries selected by the teacher, others by the student; some are same for all students, others are different. Resembles a scrapbook, providing evidence but not judging the quality of activities.
>
> The evaluation portfolio---generally standardized with substantial direction from the teacher, administator, or district. Artifacts are generally authentic and collected over time, but most entries are predetermined as are criteria for scoring and evaluating performance. There is some room for self-selection and reflection but a substantial core of required activities dominates the portfolio. (p. 337)

BEST COPY AVAILABLE

As Valencia and Calfee (1991) note, "Contrasts among the models are not trivial; they entail differing methods, criteria, purposes, and audiences." Obviously, it would be possible to construct a different typology for portfolios. However, whatever typology one wishes to employ, it is imperative that the intended purpose(s) of a portfolio assessment system be clearly specified in order to judge the sucess of the system in meeting the needs it was designed to fill.

SOME CLAIMS AND QUESTIONS

Several claims are typically made by proponents of portfolio assessment. In this section I will list the major claims that are put forth and raise basic questions with respect to these claims.

Claim 1: Portfolios are more "authentic" than other forms of assessment.

It is easy to understand the arguement that samples of the student's own work are more "meaningful and realistic" indicators of achievement than are scores on teacher designed paper and pencil tests. This is especially true in areas where the ultimate goal of instruction is to have the student produce an original piece of work, i.e. in classes on composition and in the visual arts. Just as professional writers and artists often construct portfolios of their work, it seems "natural and appropriate" to expect students to do the same. It is also logical that the portfolio would provide a student withame for all students with a sense of "ownership" over their work than do more conventional forms of assessment.

However, it seems inappropriate to assume that the portfolio concept applies equally to all areas of the curriculum and to all types of outcomes within a particular class. For example, it is not obvious that portfolios capture the intended outcomes in fields such as mathematics, science, speech, and the performing arts. Indeed, it seems quite artificial to expect students to collect sets of math, chemistry, or physics problems which they have successfully solved. It would appear that participation in science fairs or in special math competitions would be much more akin to what professionals in those fields actually do. How is this to be captured in a portfolio? Likewise, forensic competitions, drama festivals, and musical performances provide the most meaningful forums for judging accomplishments in speech and the performing arts. How does one represent such activities in a portfolio?

A more basic problem with the portfolio approach is a bias in favor of performances over more basic educational outcomes, i.e. acquiring fundamental knowledge and the comprehension of principles and/or concepts basic to a field of study. Of course, it is precisely these outcomes for which conventional paper and pencil forms of assessment are best suited. It is fashionable to argue that educators should focus upon "higher order thinking skills" instead of basic knowledge and comprehension. Those who are most adamant about the use of performance measures and portfolios make a serious mistake in assuming that newer forms of assessment should replace conventional methods. There is clearly a need for some mix of the two. However, the relative emphasis given to each approach should vary depending upon the educational outcomes deemed to be most important in a particular classroom.

Claim 2: Portfolios provide a sound basis for classroom assessment.

Any assessment procedure that is to be employed as a basis for making decisions about a student's acheivement of educational objectives should meet certain basic pyschometric standards. Advocates of assessment reform seldom give any attention to what many regard

as "technical" issues in assessment. However, the validity and reliability of asessment procedures is hardly an esoteric concern that can be ignored. As Messick (1994) has stated,

> Hence, performance assessments must be evaluated by the same validity criteria, both evidential and consequential, as are other assessments. Indeed, such basic assessment issues as validity, reliability, comparability, and fairness need to be uniformly addressed for all assessments because they are not just measurement principles, they are social values that have meaning and force outside of measurement wherever evaluative judgements and decisions are made. (p. 13)

Advocates of portfolios typically rely heavily on a"face validity" arguement in making the case that portfolios assess important outcomes of instruction. The arguement made is that portfolios are valid because they contain samples of "authentic" student work. However, the basic question posed by Arter and Spandel (1992) is,

> What is meant by authentic? The content of the portfolio will mirror the emphasis in the curriculum and classroom. For example, if the curriculum emphasizes phonics and the teachers concentrate on phonics , then the samples of work for the portfolio are likely to reflect phonics. Is this authentic? An authentic reflection of classroom work or an authentic representation of ability to read in real life? One must come to grips with this issue before even beginning to discuss authentic tasks. (p.38)

More generally, one needs to know if the work in the portfolio is representative of what a student has achieved, if the criteria used to judge performance reflect the most salient features and dimensions of the tasks, and if the conclusions drawn from the portfolio are largely a function of the individual doing the evaluation.

Few systematic studies of the psychometric properties of portfolios have been conducted. What is known is not encouraging. One of the most widely cited large-scale portfolio assessment projects is the Vermont statewide program. Klein, et.al. (1995) summarize an analysis of the reliability of the Vermont mathematics portfolio scores as follows:

> During the first 2 years of this program, there was only moderate agreement among readers regarding a portfolio's quality. This quality often varied across the five to seven pieces in a student's portfolio. Consequently, scores were not reliable enough to permit reporting student-level results. Increasing the number of readers per piece or the number of pieces per portfolio are not operationally feasible solutions to the score relibility problem. (p.243)

Given the rather extnsive resources that went into this statewide system, it seems reasonable to question whether classroom teachers without such resources should be expected to design pyschometrically sound portfolios. Given these considerations, it is difficult to regard "showcase" and "documentation" portfolios (see previous section) as serious assessment devices.

Claim 3: Portfolios provide a basis for assessing the growth of individual students.

A portfolio that is a highly personalized collection of products produced by a student naturally is likely to have substantial interest to the student and his/her parents. Likewise, samples of the student's work collected over an extended period of time provide an interesting basis for judging changes in certain skills over time. However, the judgment of "growth" is frought with problems.

BEST COPY AVAILABLE

First, exactly what is the basis for the judgment of "growth"? To be certain that one is actually assessing "growth", it is necessary to define a common scale or set of rubrics that can be validly used at different points in time. Are the performances and products collected early in the year assessing the same outcomes as those collected later in the year? Are there differences among the performances at different points in time with respect to difficulty?

Second, unless there is some standardization of portfolio content and scoring rubrics, it is impossible to make assessments of growth that generalize across students. Therefore, the more individualized the portfolios, the less meaning the notion of growth.

Third, the concept of "growth" is inherently biased in favor of the student who has the lowest level of initial achievement. It is easy to demonstrate statistically that "growth" (defined as change over time) is strongly negatively correlated with initial performance. Consequently, it is difficult to judge the "growth" demonstrated in a student's performance over time without considering the student's initial level of performance.

Claim 4: Portfolios are a practical means of classroom assessment.

Probably the most serious disadvantage of portfolios is the extreme amount of time and effort needed to implement a portfolio assessment system. Despite the assurances of advocates of alternative assessment that portfolios are a viable substitute for more traditional approaches to assessment, the need to develop scoring rubrics, rating scales, and related indices for transforming performance into quantitative results is typically glossed over.

It is absolutely essential that careful thought be given to criteria for judging the merit of individual entries in the portfolio as well as the portfolio as a whole. Further, it is necessary that the criteria be clear and publically stated in written form. As Arter and Spandel (1992) state the issue,

> Are the criteria fully and carefully defined and open to all or are they nebulous and guarded so that students must guess what is being sought? --- In the abscence of all criteria, how do we know what sort of work a student has accomplished through the year? How does the student know whether to be satisfied, ecstatic, or dismayed?

It is a monumental task to develop criteria which are agreed to by teachers with a variety of perspectives on complex performances contained in a portfolio. It is also very difficult to design scoring rubrics, rating scales, checklists, etc. that teachers can employ with a high level of reliability. Perhaps this is why few, if any, articles advocating portfolios present data regarding the inter-rater agreement or generalizability of scores across different elements contained within portfolios.

Under the heading "Staff Development", Arter and Spandel (1992) comment upon the training and skills needed by teachers in order to implement a portfolio assessment system. Among other prerequisites, they note that teachers need to:

> be well-grounded in the development and use of performance criteria so that they can recognize strong performance in writing, reading, science, math---or any area---
>
> have a great deal of content expertise so they can develop good criteria and know what to expect from students at various grade levels

be knowledgeable in the area of assessment so they can avoid the pitfalls mentioned in previous sections of this module.

To say the least, these are quite ambitious expectations. Few of the students I encounter in my classes on assessment possess these skills. Despite the efforts of myself and my colleagues to assist them in developing such skills, it seems unrealistic to expect preservive novices to acquire such sophistication in the space of one academic quarter. It is possible that some of these students will develop a high level of expertise in assessment through years of teaching experience. However, there is no guarantee that this will be the case for most teachers. It seems more likely that the majority of teachers will expend inordinate amounts of time and energy on portfolio systems wich will yield assessments of marginal psychometric value.

CONCLUSION

The notion that portfolios should be widely adopted as the primary means of classroom assessment is based upon highly questionable claims made by advocates of "school reform". Specifically, I have argued that portfolios: a. are <u>not</u> necessarily more "authentic" than traditional forms of assessment, b. do <u>not</u> offer a psychometrically sound basis for assessment, c. are based upon a flawed concept of assessing "growth", and d. require an inordinate investment of time and effort on the part of teachers.

Although portfolios may have value as instructional aids, it is not clear that they are useful assessment devices in all subject matter fields and at all grade levels. Therefore, it seems unwise to require all teachers to employ portfolios in assessing student achievement. Rather, it seems more appropriate to help teachers who wish to explore the use of portfolios for both instructional and assessment purposes. However, teachers need to be educated to become more sensitive to both the limitations and advantages of portfolios as assessment devices.

REFERENCES

Arter, J.A. and Spandel, V. (1992). Using portfolios of student work in instruction and assessment. Educational Measurement: Issues and Practice. (Spring), 36-44.

Ballard, Leslie. (February 1992). Portfolios and self-assessment. *English Journal*, 91 (2), pp. 46-48.

Farr, Roger. (1991). Portfolies: Assessment in language arts. *ERIC Digest*. Bloomington, Indiana: ERIC Clearinghouse on Reading and Communication Skills. ERIC document ED334603

Farr, Roger. (1990). Setting directions for language arts portfolios. *Educational Leadership 48* (3), p. 103.

While identifying the strengths of portfolio assessment, this article recommends careful consideration before assuming that they replacd standardized tests.

Farr, R. and Tone, B. (1994). Portfolios and Performance Assessment. Helping Students Evaluate Their Progress as Readers and Writers, Fort Worth, TX: Harcourt Brace and Co.

Gentile, Claudia. (April 1992). *Exploring new methods for collecting students' school -based writing*. NAEP's 1900 Portfolio Study. Washington, D.C.: Prepared by Educational Testing Service for the National Assessment of Educational Progress 1990 Writing Assessment for National Center of Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education. (Available from the U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, D.C. 20402-9328.)

Hansen, Jane. (April 1992). Literacy portfolios emerge. *The Reading Teacher, 45* (8), pp. 604-607.

Herter, Roberta J. (January 1991). Writing portfolios: Alternatives to testing. *English Journal, 80* (1), pp. 90-91.

Howard, Kathryn. (Spring 1990). Making the writing portfolio real. *The Quarterly of the National Writing Project and the Center for the Study of Writing and Literacy, 12* (2), pp. 4-7, 27.

Klein, S., McCaffrey, D., Stecher, B. and Koretz, D. (1995). The reliability of mathematics portfolio scores: Lessons fro mthe Vermont experience. Applied Measurement in Education, 8(3), 243-260.

Krest, Margie. (February 1990). Adapting the portfolio to meet student needs. *English Journal, 79* (2), pp. 29-34.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. Educational Researcher, 23(2), 13-23.

Murphy, Sandra, and Smith, Mary Ann. (1990). Talking about portfolios. *The Quarterly of the National Writing Project and the Center for the Study of Writing and Literacy, 12* (2), pp. 1-3, 24-27.

BEST COPY AVAILABLE

NEA, (1993), Student Portfolios, An NEA Professional Library Publication, Teacher-to -Teacher Series, Washington, National Education Association.

Paulson, L., Paulson, P., and Meyer, C. (1991). What makes a portfolio a portfolio? Educational Leadership, 48(5), 60-63.

Portfolios illuminate the path for dynamic, interactive readers. (May 1990). (Secondary Perspectives.) *Journal of Reading, 33* (8), pp. 644-647.

Russavage, Patricia McGrath. (1992). Building credibility for portfolio assessment. In *Literacy: Issues and practices.* 1992 Yearbook of the State of Maryland International Reading Association Council, Vol. 9.

Sustein, Bonnie S. (1992). The personal portfolio: Redefining literacy, rethinking assessment, reexaming evaluation. *Writing Notebook: Creative Word Processing in the Classroom*, 9 (4) pp. 36-39.

Tierney, Robert J., Carter, Mark A., and Desai, Laura E. (1991). *Portfolio assessment in he reading-writing classroom.* Norwood, Massacusetts: Christopher-Gordon.

Valencia, S. (1990). A Portfolio approach to classroom reading assessment: the whys, whats, and hows. The Reading Teacher, 43(4), 338-340.

Valencia, S. and Calfee, R. (1991). The development and use of literacy portfolios for students, classes, and teachers. Applied Measurement in Education, 4(4), 333-345.

Yancey, Kathleen Blake (ed.) (1992). *Portfolios in the writing classroom: An introduction.* Urban, Illinois: National Council of Teachers of English.